IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2024

DOI: 10.17148/IJIREEICE.2024.12524

Online Fraudulent Transaction Detection Through Machine Learning and Deep Learning Algorithms

Prof. Nikita P. Shah[1], Komal Balaji Panchal[2], Vaishnavi Ashok Jambhale[3], Gauri Kaluram Kharat[4], Siddhi Narendra Galinde[5]

Dept. of Computer Engineering VPKBIET, Baramati, India[1-5]

Abstract: The virtual world has led to a rise in credit card use in the modern era, but misuse and fraud of credit cards have also increased dramatically. It is necessary to identify the many kinds of credit card fraud. Such frauds cause significant financial losses for both the business and the cardholder. Determining whether or not a specific transaction is fraudulent is the primary goal. A high false alarm rate, a shift in the nature of fraud, access to public data, and a large class imbalance are all necessary for detecting fraud. It acknowledges the challenges posed by imbalanced data and explores a range of machine learning and deep learning algorithms. The study focuses on convolutional neural networks (CNNs) and their architectural variations to enhance fraud detection. Through empirical analysis, it achieves impressive results, outperforming existing methods with high accuracy, F1-score, precision, and AUC values. The research also emphasizes the importance of minimizing false negatives. Ultimately, the proposed deep learning model offers a promising solution for real-world credit card fraud detection.

Index Terms: credit card fraud, machine learning, deep learning, CNN

I. INTRODUCTION

Since credit cards are a convenient and effective tool for online transactions, most of the people are using them. However, this has also increased the potential for credit card misuse. Both credit card providers and cardholders suffer large financial losses as a result of credit card theft.

The rise in online purchasing has also led to an increase in card-not- present fraud, or the use of your credit card number in online transactions. As fraudsters are always coming up with new techniques, fraud detection systems need to be updated on a regular basis. Since the security of a system depends on the protection of several layers, each component should be sold separately. Because credit card transaction datasets are extremely uncommon and highly skewed, preprocessing is necessary before using them. Assembling techniques can be Identify applicable funding agency here. If none, delete this.

utilized to lessen variance and produce accurate and useful outcomes. It's fairly simple to do a credit card transaction, but since fraudsters usually attempt to pass off their illegal activities as legitimate, detecting fraud gets more challenging. When someone other than the owner uses a credit card or account data to make an unauthorized transaction, it's known as credit card fraud. Fraud may be committed with a lost, stolen, or counterfeit credit card.

The rise in online purchasing has also led to an increase in card-not-present fraud, or the use of your credit card number in online transactions. This occurs when stolen credit card information is used to make purchases online, bypassing traditional verification methods. To address these challenges, fraud detection systems must continuously update to stay ahead of fraudsters' evolving techniques. These systems rely on multiple layers of security measures, each independently robust, to effectively identify and prevent fraudulent transactions.

However, analyzing credit card transaction data presents its own challenges. Datasets are often rare and skewed, requiring preprocessing to ensure accurate analysis. Techniques like data assembly can help reduce variance and improve the reliability of insights. Detecting fraud is further complicated by fraud- sters' efforts to disguise their activities as legitimate transac- tions. Credit card fraud occurs when unauthorized individuals use stolen or counterfeit card information for purchases, with online transactions providing fertile ground for such activities.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2024

DOI: 10.17148/IJIREEICE.2024.12524

II. LITERATURE SURVEY

A. Credit card fraud detection using artificial neural network

Nowadays, as most of us use credit cards more regularly for payments, credit card fraud is a typical occurrence. This is because of the development of technology and the rise in online transactions, which lead to frauds that cause enormous financial losses. As a result, efficient techniques to lessen the loss are required. This work attempts to forecast the occurrence of fraud by utilizing several machine learning algorithms, including support vector machines (SVM), k-nearest neighbors (Knn), and artificial neural networks (ANNs). To distinguish between transactions that involve fraud and those that don't, it involves applying sophisticated supervised machine learn- ing and deep learning algorithms. Artificial neural networks (ANNs) are used, and they provide enhanced accuracy which is most suitable for identifying credit card fraud.It provides greater accuracy than algorithms for unsupervised learning. Pre-processing, normalization, and undersampling were used in this dataset to address issues arising from the use of an unbalanced dataset.

B. An efficient real time model for credit card fraud detection based on deep learning

This paper presents a dual-focused investigation into credit card fraud detection. Firstly, it introduces a real-time deep learning approach utilizing autoencoders. The study prioritizes the real-time classification aspect, employing a prediction model, deep autoencoder, and deep neural network. Sec- ondly, the paper conducts a thorough comparison of binary classification methods within the domain of financial fraud, encompassing regression models, SVM, restricted Boltzmann machines, and ANN, with a specific emphasis on logistic regression. The research evaluates these methods using a dataset of European card transactions over a two-day period in September 2012. Through rigorous testing and construction of Confusion Matrices, the study reveals the performance metrics of each algorithm. Notably, the deep neural network-based autoencoder outperforms other models with an F1 score of 0.294, contributing valuable insights to the ongoing discourse in credit card fraud detection and aiding in the development of robust security measures in the financial sector.

C. Malware classification with improved convolutional neural network mode

The study emphasizes the enhancement of CNN models through integrated data pre-processing and augmentation techniques, accommodating computing resource constraints. The CNN model presented in the research attains a noteworthy accuracy of 98.03 percent. A subsequent advancement involves the introduction of a hybrid CNN+L2-SVM model, replacing the Softmax activation function with the recommendation of the Support Vector Machine. This modification yields a remarkable accuracy of 99.59 percent, outperforming the Softmax approach by 1.56 percent. The hybrid model not only demonstrates superior accuracy but also proves effective in misidentifying malware samples. Its efficiency, requiring reduced time and computing power, is attributed to a compre- hensive strategy that addresses data imbalance, incorporates image scaling, fine-tunes the CNN model, employs multi- class SVM, and strategically replaces Softmax with L2-SVM. This approach, leveraging squared slack variables in L2-SVM, contributes to more broadly applicable findings.

III. RESEARCH METHODOLOGY

A. Description of Dataset

The European card dataset utilized in this study consists of 284,807 transaction records, forming a vital dataset for credit card fraud detection systems. Each transaction entry contains detailed information, including timestamps indicating transac- tion times, transaction amounts representing monetary values, and merchant category codes (MCCs) classifying transactions by merchant types. Additionally, anonymized cardholder data, including demographic details and transaction histories, is included for analysis of cardholder behaviors. To address the dataset's high dimensionality and enhance computational efficiency, Principal Component Analysis (PCA) is performed on selected features. Furthermore, strict measures are taken to ensure compliance with data protection regulations, such as the General Data Protection Regulation (GDPR), through anonymization techniques that protect cardholders' privacy while maintaining data integrity.

B. Data preprocessing

In the preparation of the European card dataset for analysis within credit card fraud detection systems, a series of steps are implemented to ensure data quality and suitability for subsequent modeling. Initial steps involve addressing miss-ing values, employing techniques like imputation or removal to maintain dataset integrity. Subsequently, outlier detection methods are applied to identify and manage anomalies that could affect analysis outcomes. Following this, normalization techniques are employed to standardize feature scales, promot- ing consistency and enhancing the performance of machine learning algorithms.

DOI: 10.17148/IJIREEICE.2024.12524

Categorical variables, such as merchant category codes (MCCs), are encoded using strategies like one- hot encoding to facilitate their integration into predictive models. Furthermore, to mitigate the impact of imbalanced class distribution between fraudulent and non-fraudulent trans- actions, sampling techniques such as oversampling or under- sampling may be utilized to achieve a more balanced represen- tation. These preprocessing measures collectively contribute to optimizing the European card dataset for subsequent credit card fraud detection analysis.

C. Feature Engineering

To enhance the European card dataset for fraud detection, Principal Component Analysis (PCA) is employed to effec- tively reduce dimensionality while preserving key information. Initially, relevant features such as timestamps, transaction amounts, and merchant category codes (MCCs) are extracted from the raw transactional data. Subsequently, PCA is ap- plied to transform these features into a set of orthogonal components, denoted as $v1$, $v2$, $v3$, ..., $v28$. These components capture different patterns of variance within the dataset, with higher-order components capturing increasingly finer details. By retaining the most significant components explaining the majority of the variance, PCA effectively reduces dimension- ality while retaining essential information critical for fraud detection analysis. These transformed features, represented by $v1$ through $v28$, serve as input for subsequent machine learning algorithms, facilitating more efficient model training and inference without compromising predictive performance.

D. Performance Evaluation Measures

In assessing the effectiveness of credit card fraud detection systems on the dataset, several performance evaluation mea- sures are employed to gauge model performance accurately. These measures include:

1) Accuracy: The proportion of correctly classified trans- actions (both fraudulent and non-fraudulent) relative to the total number of transactions in the dataset. While accuracy provides a general overview of model performance, it may be misleading in the presence of imbalanced datasets.

2) Precision: The proportion of correctly identified fraud- ulent transactions among all transactions predicted as fraud- ulent. Precision measures the model's ability to avoid false positives, indicating the proportion of flagged transactions that are genuinely fraudulent.

3) F1-Score: The harmonic mean of precision and recall, providing a balanced measure of a model's performance. F1- score considers both false positives and false negatives, offering a single metric to assess overall model effectiveness.

4) Recall: The proportion of correctly identified fraudulent transactions among all actual fraudulent transactions in the dataset. Recall measures the model's ability to capture all instances of fraud, indicating the proportion of fraudulent transactions correctly identified by the model.

IV. ALGORITHMS

Credit card fraud detection is a critical application of machine learning and data analysis to identify potentially fraudulent transactions and protect both consumers and finan- cial institutions from unauthorized charges. Various algorithms and techniques can be used for this purpose, including logistic regression, support vector machines (SVM), and convolutional neural networks (CNNs).

A. Logistic Regression

Logistic regression is a well-known classification algorithm extensively used for binary classification tasks, such as the detection of fraud. Its primary objective is to categorize trans- actions as either legitimate or fraudulent, and one of its key advantages is its interpretability, which enables a clear under- standing of the factors influencing classification decisions. At its core, logistic regression serves as a statistical and machine learning tool specialized for binary classification challenges.

It is employed to predict one of two possible outcomes based on input features. This methodology is particularly effective when the target variable is binary and categorical, encompassing values like yes/no, true/false, or 0/1. The process of training logistic regression models involves using labeled data, where legitimate transactions are designated as "0," and fraudulent transactions are marked as "1." Through training, logistic regression establishes a decision boundary within the feature space, effectively segregating the two classes. The decision boundary may take on linear or nonlinear forms, depending on the relationship between input features and the binary outcome.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2024

DOI: 10.17148/IJIREEICE.2024.12524

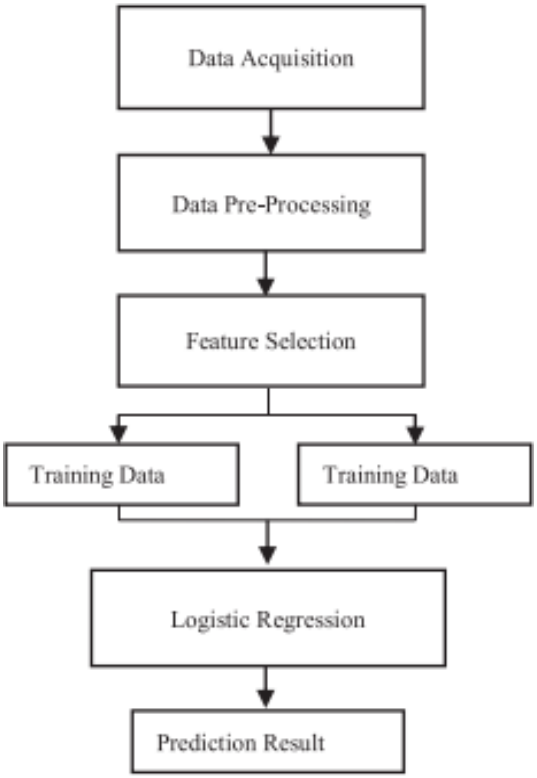


Fig. 1. Flowchart

B. Support Vector Machine

Support Vector Machines (SVM) are a valuable machine learning method applicable to credit card fraud detection, espe- cially when confronted with complex and non-linear datasets. The task of credit card fraud detection involves binary classi- fication, where the aim is to differentiate between legitimate and fraudulent credit card transactions. SVM can be utilized to establish an effective decision boundary that maximizes the distinction between these two classes. To ensure uniform scales for features and prevent certain features from unduly influencing the decision-making process, feature scaling is often applied to the data before using SVM.

SVM's primary goal is to identify the hyperplane that max- imizes the margin between the two classes. This hyperplane, referred to as the "support vector," plays a crucial role in achieving the most effective separation between legitimate and fraudulent transactions. SVM is a robust tool in the field of credit card fraud detection, particularly when dealing with in- tricate data and non-linear relationships. It has the capability to create an optimal boundary that effectively separates legitimate and fraudulent transactions and can be fine-tuned for peak performance.

C. CNN

For the credit card fraud detection system, a Convolutional Neural Network (CNN) model is implemented due to its capacity to capture intricate patterns and relationships within transaction data. The implementation involves several key steps:

1) *Data Preparation::* The European card dataset, com- prising transaction records, undergoes preprocessing to extract pertinent features and ready the data for model training. This includes addressing missing values, detecting outliers, and normalizing the data to ensure consistency and reliability.

2) *Feature Engineering::* Significant features related to transaction timestamps, amounts, and merchant category codes (MCCs) are chosen and engineered to provide meaningful rep- resentations of transaction data. Additional features capturing temporal patterns or cardholder behaviors may be created to enhance model performance.

3) *Model Architecture::* The CNN model architecture is designed to accommodate the sequential nature of transac- tion data. It typically comprises multiple convolutional layers followed by pooling layers to extract spatial features and reduce dimensionality. Optionally, recurrent layers like Long Short-Term Memory (LSTM) cells may be included to capture temporal dependencies within transaction sequences.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2024

DOI: 10.17148/IJIREEICE.2024.12524

4) *Model Training::* The prepared dataset is split into training and validation sets, and the CNN model is trained using the training data. During training, the model learns to classify transactions as fraudulent or legitimate based on provided features. Optimization techniques such as gradient descent and backpropagation are employed to update model parameters and minimize the loss function.

5) *Model Evaluation::* The trained CNN model is evaluated using a separate test set to assess its performance in detecting fraudulent transactions. Metrics such as accuracy, precision, recall, F1-score, and ROC-AUC are computed to quantify the model's effectiveness in identifying fraudulent activities while minimizing false positives.

6) *Fine-Tuning and Optimization::* Further fine-tuning and optimization of the CNN model may be performed to en- hance performance. This could involve hyperparameter tuning, adjusting model architecture, or incorporating regularization techniques to prevent overfitting and improve generalization capabilities.

7) *Deployment and Monitoring::* Upon demonstrating sat- isfactory performance, the CNN model is deployed into the credit card fraud detection system for real-time monitoring of transaction activities. Continuous monitoring and periodic retraining ensure the model's effectiveness in detecting evolv- ing fraud patterns and maintaining robustness against emerg- ing threats. Implementing a CNN model tailored for credit card fraud detection enables financial institutions to leverage advanced deep learning techniques to enhance accuracy and protect cardholders from fraudulent activities effectively.

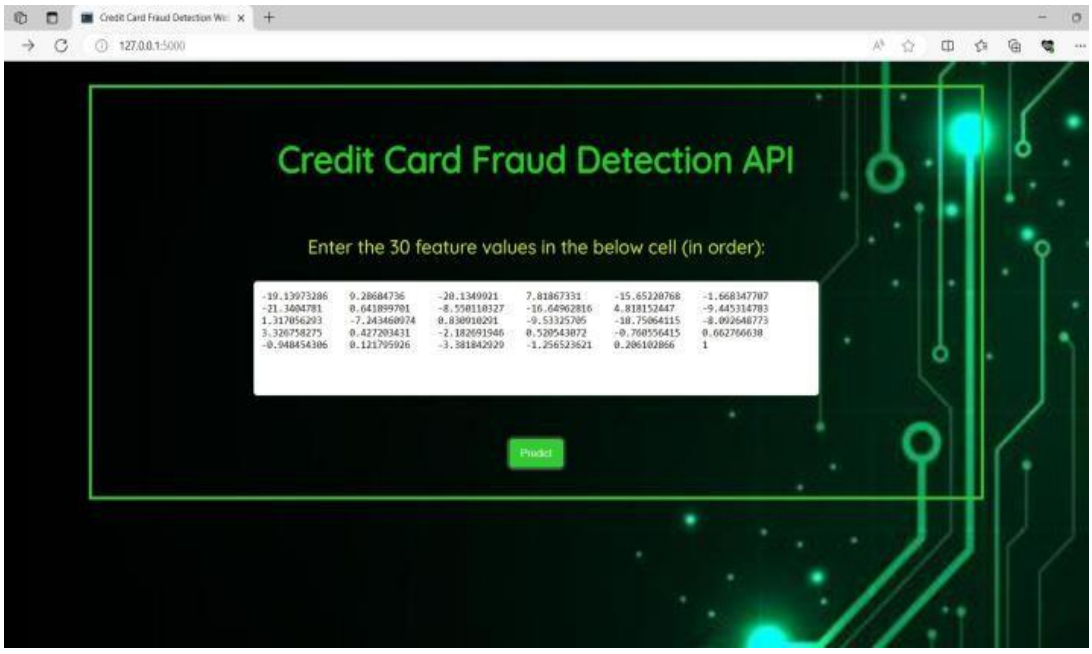
V. RESULTS AND DISCUSSIONS

A. Presentation of the Experimental Setup:

To evaluate credit card fraud detection on the European card dataset, the dataset was divided into training, validation, and test sets. Various machine learning and deep learning algo- rithms, including logistic regression, decision trees, random forests, support vector machines (SVMs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs), were utilized to construct fraud detection models. Feature engineering techniques were applied for extracting relevant features, while preprocessing steps such as normalization and dimensionality reduction were performed to enhance model performance. Model evaluation utilized metrics like accuracy, precision, recall, F1-score, AUC-ROC, and AUC-PR.

B. Results of Applying Machine Learning and Deep Learning Techniques:

Across different models, varying results were observed. While logistic regression and decision trees showed moderate performance, random forests and SVMs displayed improved results, especially in addressing class imbalance. Deep learn- ing techniques, including CNNs and RNNs, exhibited promis- ing outcomes by capturing intricate patterns within transaction data, resulting in enhanced fraud detection accuracy.

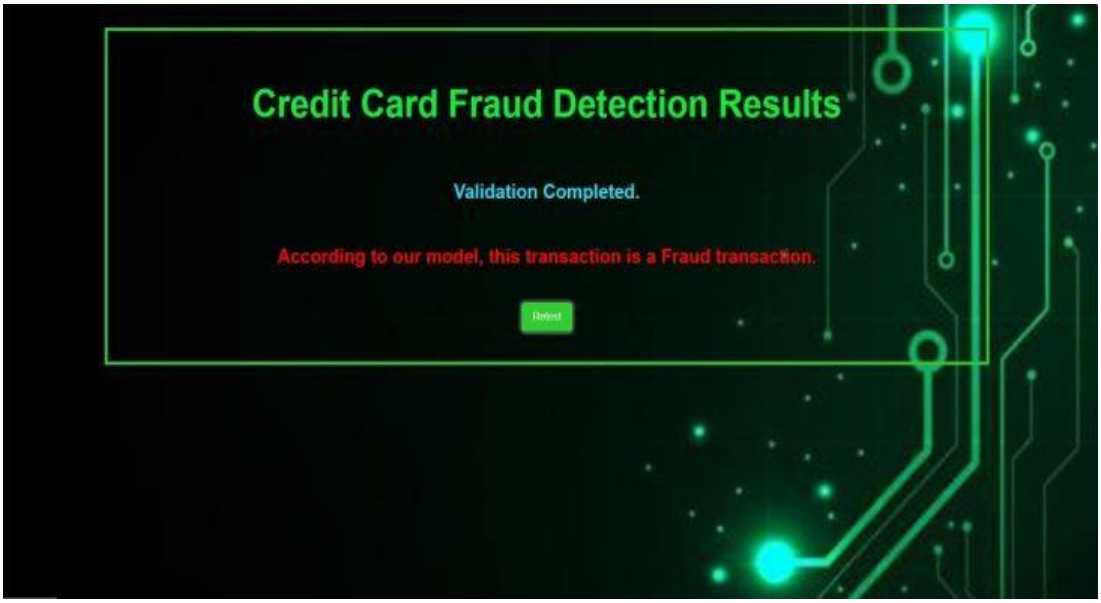


IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2024

DOI: 10.17148/IJIREEICE.2024.12524



C. Comparison with Baseline Methods or Existing Ap- proaches:

The developed models outperformed baseline methods and existing approaches. Traditional machine learning algorithms served as benchmarks, highlighting the effectiveness of ad- vanced techniques like random forests and deep learning architectures. This comparison underscored the significance of leveraging sophisticated algorithms and feature engineering strategies to enhance fraud detection accuracy and tackle challenges associated with imbalanced datasets.

D. Discussion of Performance Metrics and Findings:

The analysis of performance metrics and findings shed light on the strengths and limitations of different fraud detec- tion approaches. While machine learning algorithms offered interpretability and competitive performance, deep learning techniques excelled in capturing complex patterns in trans- action data. Evaluation of performance metrics emphasized the trade-offs between precision, recall, and overall model accuracy, emphasizing the need for a balanced approach in fraud detection. Furthermore, potential areas for improvement, such as refining feature engineering and exploring ensemble learning techniques, were discussed to further enhance fraud detection performance. Overall, the experimental results pro- vided valuable insights into the efficacy of machine learning and deep learning techniques for credit card fraud detection on the European card dataset, laying the groundwork for more effective fraud prevention measures.

VI. CONCLUSION

In this study, our exploration centered on evaluating the efficacy of Logistic Regression (LR) and Convolutional Neu- ral Network (CNN) techniques for the detection of online fraudulent transactions. Through our experimentation, we un- covered promising results for both LR and CNN models, showcasing their potential in effectively identifying fraudulent activities within online transactions. LR, recognized for its simplicity and interpretability, served as a robust baseline, exhibiting commendable accuracy in discerning fraudulent behavior. Conversely, the CNN model, renowned for its capac- ity to capture intricate patterns embedded within transaction data, emerged as a standout performer, particularly adept at identifying nuanced and complex fraud indicators.

The amalgamation of LR and CNN techniques presents an enticing opportunity for financial institutions to bolster their fraud detection capabilities, thereby mitigating the risks associated with online fraudulent transactions. Moving forward, our study aims to delve deeper into the interpretability of CNN models and explore innovative feature engineering strategies to further enhance the performance and robustness of fraud detection systems. Additionally, we envision the integration of real- time monitoring mechanisms and the exploration of ensemble learning techniques to fortify the resilience of fraud detection frameworks against evolving fraudulent tactics. By continually refining and advancing our methodologies, we strive to stay at the forefront of combating online fraud, safeguarding the integrity of financial transactions and bolstering consumer trust in the digital realm.

DOI: 10.17148/IJIREEICE.2024.12524

REFERENCES

[1]. A. Rb and S. K. Kr, ”Credit card fraud detection using artificial neural network”, Global Transitions Proc., vol. 2, no. 1, pp. 35-41, Jun. 2021.

[2]. Y. Abakarim, M. Lahby and A. Attioui, ”An efficient real time model for credit card fraud detection based on deep learning”, Proc. 12th Int. Conf. Intell. Systems: Theories Appl., pp. 1-7, Oct. 2018.

[3]. S. S. Lad, I. and A. C. Adamuthe, ”Malware classification with improved convolutional neural network model”, Int. J. Comput. Netw. Inf. Secur., vol. 12, no. 6, pp. 30-43, Dec. 2021.

[4]. H. Najadat, O. Altiti, A. A. Aqouleh and M. Younes, ”Credit card fraud detection based on machine and deep learning”, Proc. 11th Int. Conf. Inf. Commun. Syst. (ICICS), pp. 204-208, Apr. 2020.

[5]. P. Raghavan and N. E. Gayar, ”Fraud detection using machine learning and deep learning”, Proc. Int. Conf. Comput. Intell. Knowl. Economy (ICCIKE), pp. 334-339, Dec. 2019.

[6]. D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic and A. Anderla, ”Credit card fraud detection–machine learning methods”, Proc. 18th Int. Symp. INFOTEH-JAHORINA (INFOTEH), pp. 1-5, Mar. 2019.

[7]. X. Zhang, Y. Han, W. Xu and Q. Wang, ”HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture”, Inf. Sci., vol. 557, pp. 302-316, May 2021.

[8]. S. Warghade, S. Desai and V. Patil, ”Credit card fraud detection from imbalanced dataset using machine learning algorithm”, Int. J. Comput. Trends Technol., vol. 68, no. 3, pp. 22-28, Mar. 2020.

[9]. N. Kousika, G. Vishali, S. Sunandhana and M. A. Vijay, ”Machine learning based fraud analysis and detection system”, J. Phys. Conf., vol. 1916, no. 1, May 2021.

[10]. J. Forough and S. Momtazi, ”Ensemble of deep sequential models for credit card fraud detection”, Appl. Soft Comput., vol. 99, Feb. 2021.